STAT 542 - Skin Cancer Diagnostics Group Project

By Pulkit Dixit (pulkitd2), Matthew Modaff (mmodaff2) and Tejo Venkat Nutalapati (nutalpt2)

Introduction:

We set out to detect and classify skin images into two categories: benign and malignant. To this end, we imported images, performed data processing (discussed in the next section) and image denoising, and the proceeded to classify the images into one of the two afore-mentioned labels and analyzed their results.

The first approach we used involved identifying malignant moles in the skin images based on their pixel values. For this we used 3 linear and 3 non-linear classifiers to perform pixel-based classification on the denoised images, and report the accuracies for the three best models along with some more metrics and analyses. The results for the pixel-based classification were not very encouraging. The classification algorithms used were K-Nearest Neighbors, Linear Discriminant Analysis (LDA), and Linear Support-vector Machines (SVM), and the best accuracies obtained for them were 0.59, 0.57 and 0.59 respectively. These results indicated that pixel-based classification for a small dataset of similar images cannot obtain reliable predictions for skin cancer images.

Our second approach involved extracting features from the images in order for our models to be more interpretable so that we may convince experts in the medical field of interesting features for identifying malignant moles. Following a literature review of various papers that focused on similar work (discussed later in the report), we generated features from both classes of images and performed classification on the new set of features. The results for this course of action produced much better results, obtaining accuracies of 0.72 and 0.70 on our test set for random forest and penalized logistic regression models.

Data Processing:

The data was placed in two separate folders - one for benign images and one for malignant images. Both folders contained 150 images each. There were no null images observed in the data import process. However, we did notice that almost all the images were of different sizes. This would create problems during classification as each image would have a different number of features. To rectify this, we resized the images to be 256x256 in height and width. While classifying the images, we experimented with retaining the 3 channels of each image and converting them to grayscale. The first option proved to be better for pixel-based classification, while grayscale images were used in some features. Post resizing, we normalized the images and flattened them into one dimensional vectors so that they could be used as variables for classification.

Apart from this, we also noticed that the images contained noise in the form of hairs and skin color variations (and possibly other factors as well) which could hamper classification results [1]. In order to alleviate this, we denoised images by performing erosion - to remove black noise - and dilation - to remove white noise - on each image. Figure 1 shows a before-after comparison of an image.

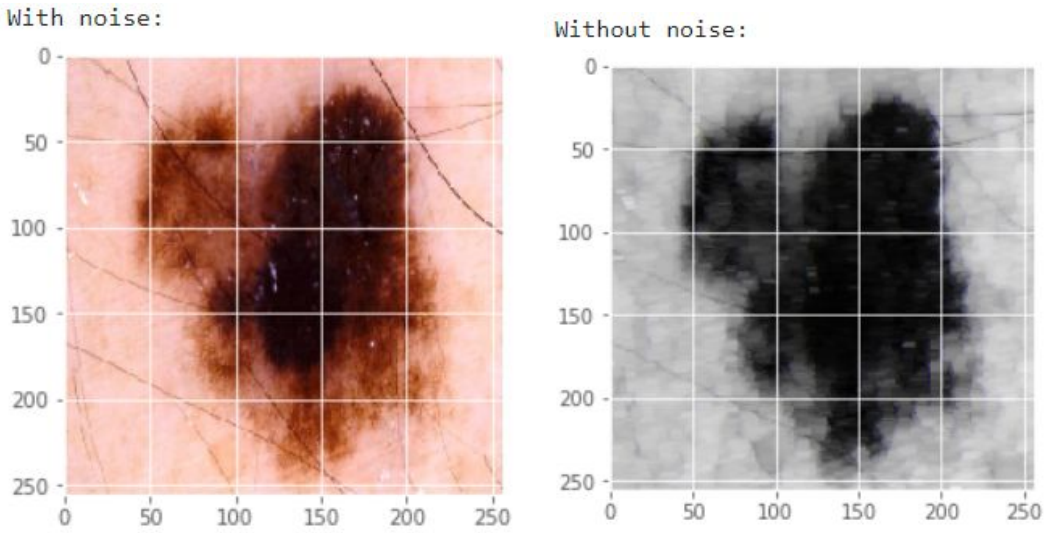


Figure 1

As can be seen from the above figure, the image on the right contains much less hair effect than the one on the right. This procedure helped us generate better accuracies for the pixel-based approaches.

Pixel-based Classification:

Dimension Reduction:

Since each image after pre-processing was 256x256x3 elements long, and there were only 300 images in total, dimension reduction was needed so that each classification model could operate on a limited set of features. For this purpose, we tested two techniques:

- Independent Component Analysis (ICA): separates data into independent components. Since the images here were RGB images, we intuitively assumed that ICA could separate the images into independent parts similar to how signal isolation is performed.
- Principal Component Analysis (PCA): separates data into variables such that they explain as much variance in the data as possible. Since this technique aims to explain more variance in fewer variables, it seemed apt to apply it to our dataset.

Classification Models:

The models we used are listed below along with the rationales based on which we selected them:

1. K-Nearest Neighbors (KNN): Naively, we would assume that the pixels of the malignant tumor images are most similar to the pixels of other malignant tumor images.
2. Linear Discriminant Analysis (LDA): LDA works specifically to separate the images into distinct classes by finding a linear combination of features that characterize each class of images. Intuitively, it makes sense to use this algorithm on the basis of capturing the possible interaction between pixels color values as well as intensities of related pixels. For example the values in the red, blue, and green color channels could hold a possible link to characterizing images in the malignant dataset and LDA would help capture this relationship
3. Linear Support-vector Machine (SVM): With the Linear SVM, we hoped to determine a hard boundary between the feature space of pixels in the malignant and benign images.

Results:

The results obtained for each model are explained below:

1. Model 1: The K-Nearest Neighbors (KNN) classifier made the most sense from the beginning with the assumption that malignant and benign images had similar pixel values within their own class. If there was a certain pattern to the pixels that was indicative of malignant or benign melanomas, the KNN classifier would hopefully pick up on it. Our KNN classifier did the best out of all of our classification models, presumably for that reason. After tuning, we elected to use 3 nearest neighbors for the ICA and 7 nearest neighbors for the PCA

   Table:

Dimension Reduction Technique	Number of components	Accuracy	Precision	Recall
ICA	3	.59	.64	.49
PCA	3	.58	.61	.53

   Confusion matrix: (ICA)

Predicted Label\True Label	Benign	Malignant
Benign	30	13
Malignant	24	23

2. Model 2: Linear Discriminant Analysis (LDA) allows us to find features from the pixels with which we can separate the class

   Table:

Dimension Reduction Technique	Number of components	Accuracy	Precision	Recall
ICA	3	0.57	0.62	0.45
PCA	3	0.57	0.62	0.45

   Confusion matrix: (ICA)

Predicted Label\True Label	Benign	Malignant
Benign	31	12
Malignant	27	20

3. Model 3: The Linear SVM made sense to use as we would expect there to be a boundary between the malignant image pixels and the benign image pixels, for instance that the malignant images might have melanomas that have a certain shape or color which would be shown through the pixels.

   Table:

Dimension Reduction Technique	Number of components	Accuracy	Precision	Recall
ICA	3	.57	0.62	0.45
PCA	3	.59	0.63	0.51

   Confusion matrix: (PCA)

Predicted Label\True Label	Benign	Malignant

Benign	29	14
Malignant	23	24

Analysis of the Failure of Pixel-based Classification:

Upon seeing un-encouraging results for all of our pixel-based models we re-ran our dimension reduction techniques for only 2 components each and plotted them so check if a good decision boundary could have been plotted by any model. Figure 2 shows the corresponding plots for ICA and PCA:

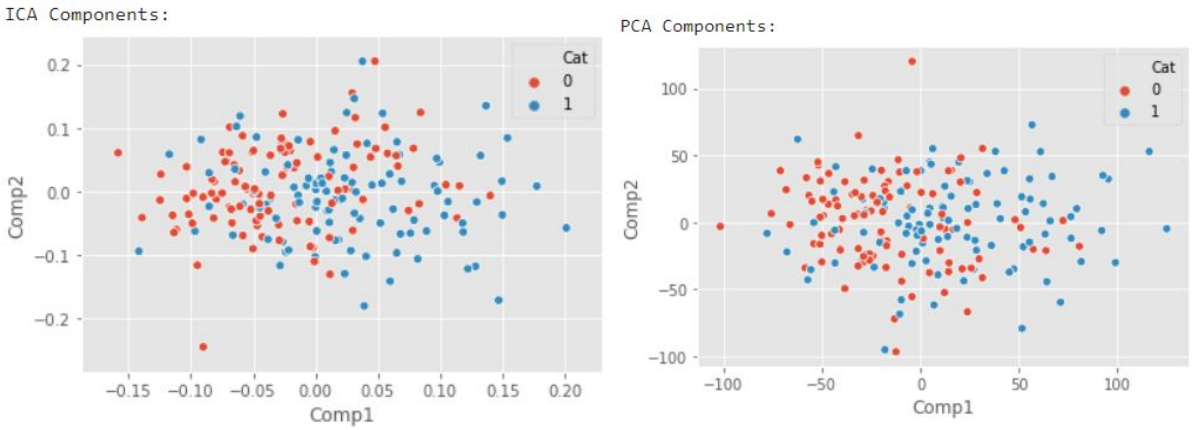


Figure 2

Both scatterplots show that there is no distinct decision boundary - linear or non-linear - between the two classes of images. This is why none of the models attain accuracies over 60%. This shows that classifying skin cancer images based on pixel values is not an appropriate way to proceed with the classifying problem. Intuitively, this makes sense since the images contain skin, moles and sometimes hairs, all of which are extremely similar in color. This would make pixel-based classification hard, especially for a dataset of only 300 images as the classification algorithms would not be able to learn the differences between the two images. As an alternate solution feature extraction (intensity, color gradient, edge detection, etc.) could be performed on the images and used for classification instead.

Feature-based Classification:

Literature Review:

Because diagnosing melanomas early increases chances of survival greatly, doctors have developed metrics used for differentiating between malignant and benign moles. Many of these

techniques can be applied by patients, and can be checked using visible cues. We can take these techniques and apply them to our data set to give us interpretable features.
The techniques follow the mnemonic of ABCDE - Asymmetry, Borders, Color, Diameter or Dark, and Elevated or Evolving [2].

- Asymmetry - a mole that is mostly oval or circular is likely benign, while one that is not mostly the same on both sides is more likely to be malignant.
- Border - If the mole has an irregular border (uneven, notched edges, etc) it is more likely to be malignant.
- Color - Melanomas can grow spots that are red, blue, or white, or has uneven colors as they grow.
- Diameter - If the mole is larger than about 6 mm in diameter, it's more likely to be malignant.
- Dark - Moles that are darker than others might be a warning sign.
- Evolving - Moles that change over time are a warning sign.
- Elevated - Moles that have a surface that is higher than the surrounding skin is more likely to be malignant.

Of those criteria, a few right off the bat are not useful for classification of images. Elevated is very difficult or impossible to tell from these images. Diameter is also out, as we do not know the scale at which all of these images were taken, or if they were even taken all at the same scale. Evolving also will not work, since we only have a snapshot at one point in time for each example.

Feature Engineering:

Based on these criteria, we developed a number of different features for classifying the pictures based on the scores each image has for each feature.

The first feature used is Asymmetry. To measure it, we took each image, padded the image into a square using the average of border pixels, then dropped a circular mask on top of it. Then, at angles of 0 to 90 degrees, incrementing by 5, we rotated the image and then compared the pixel values of the left side to its mirror on the right. We then took the mean squared difference between the red, green, and blue pixel values of the mirrored images for each rotational value, and then took the minimum. That score became our asymmetry feature score.

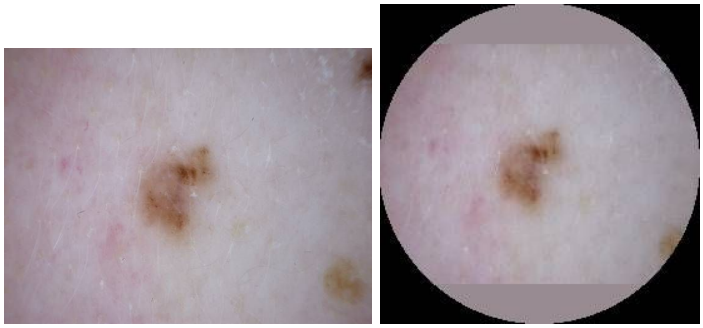


Figure 3

The second feature is testing for the irregular border. To do this, we resized each image to a standard size of 256 to 256, then ran opencv's edge detection algorithm, which spits out an image that has each pixel on or off, depending on if it's considered a border pixel or not. The score was then just an average of the pixel values in this edge detected image.

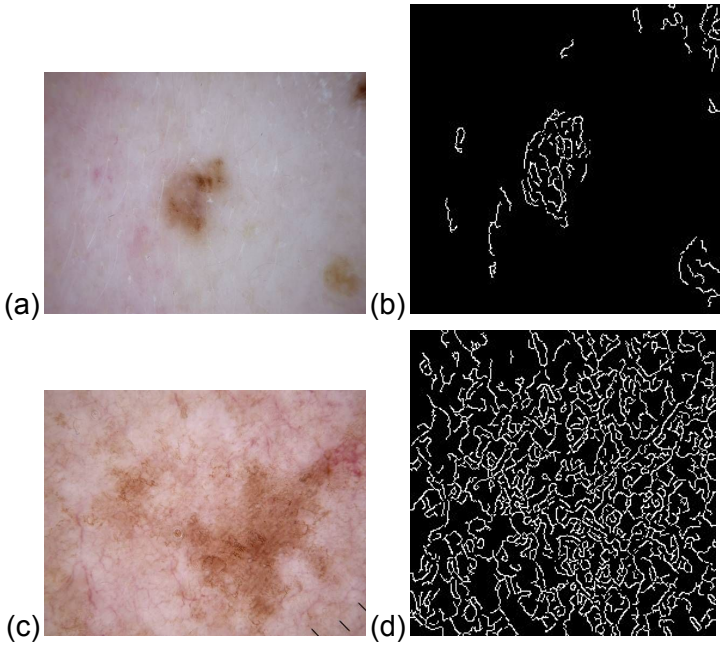
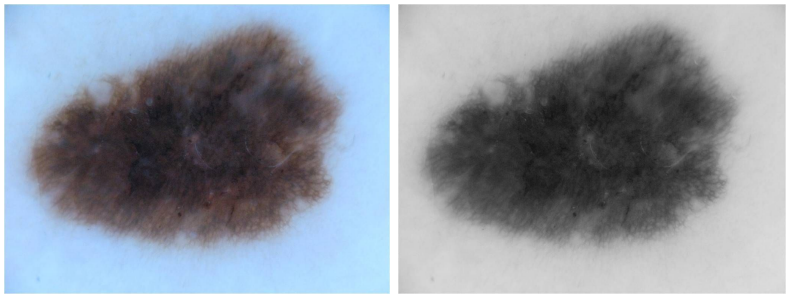
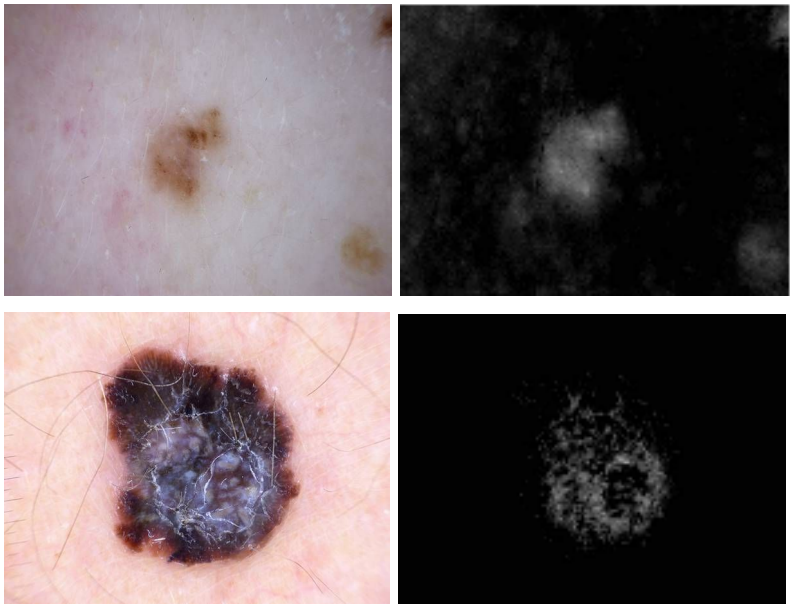


Figure 4

Here we have (a) an image from the benign dataset, (b) its edge detection results, (c) an image from the malignant dataset, and (d) its edge detection results

The last features involved analyzing the colors of each of the images. To test for dark/white, each image was converted to grayscale and then each pixel was averaged out to get the score. To test for blue, each pixel's blue channel was subtracted by the red or green channel, whichever was bigger. These values were then averaged out for one feature. This process was then repeated for the red channel.





This then gives a feature set with 5 columns, each of which is interpretable for a doctor, and follows the clinical methods for diagnosing melanomas. We then ran two different classification methods on the data set, random forest and penalized logistic regression.

Classification Models:

First, we chose to implement a penalized logistic regression on the features. This is because a logistic regression would be easier to interpret for a collaborator or doctor and much easier for them to explain to patients. The penalization in the logistic regression was to prevent overfitting.

We encountered correlation in our logistic regression so using a tree-based model would counteract that. TheIt creates a combination of more consistent accuracy values
For the Random Forest classification model, using ntree = 500, and mtry = 154 (after tuning), on our test set of 74 pictures.

Predicted Label\True Label	Benign	Malignant
Benign	28	12
Malignant	9	25

Accuracy	Precision	Recall
0.72	0.74	0.68

Variable	Variable Importance
edge	100.00
red	26.11
blue	12.03
intensity	10.49
asymmetric	0.00

For the Penalized Logistic Regression model, using lambda = 0.04145973,

Predicted Label\True Label	Benign	Malignant
Benign	29	14
Malignant	8	23

Accuracy	Precision	Recall
0.70	0.74	0.62

Variable	Variable Importance
edge	100.00
blue	34.02
intensity	27.24
asymmetric	14.56
red	0.00

Conclusion:

Going from a pixel-based approach (Part 1) to a more feature-based approach (Part 2) greatly improved our classification results. The most accurate model we had in Part 1 had an accuracy of 0.59, whereas the most accurate model in Part 2 had an accuracy of 0.72. The pixel-based

classification results told us that there was little we could tell from the pixels to determine whether a skin image was malignant or benign. This approach did a good job of analyzing the effect of intensities and color but did little to find the overall patterns within the image. The feature-based approach yielded better results as these were driven by insights from which

From the feature-based classification, we learned that the edges were the most important feature in determining the malignant images, according to the variable importance. The precision and recall along with the type 1 and type 2 errors point towards the edges being significant and not just opportunistic noise introduced into the predictions. Similarly, both models agree that edges are the most important feature in determining malignant images. This is in agreement with the clinical diagnostic criteria from [2]. Our edge detection feature proves to be important both statistically and clinically.

References:

[1][https://www.sciencedirect.com/science/article/pii/S0010482597000206?via%3Dihub]
[2][https://emedicine.medscape.com/article/1295718-overview]
[3][https://www.skincancer.org/skin-cancer-information/melanoma/melanoma-warning-signs-and-images/]